GEOMETRY-DRIVEN VISUALIZATION OF MICROSCOPIC STRUCTURES IN BIOLOGY

K. Mosaliganti, R. Machiraju

The Ohio State University
Computer Science and Engineering
2015 Neil Avenue, Columbus, OH-43210

K. Huang

The Ohio State University
Biomedical Informatics
333 W. 10th Avenue, Columbus, OH-43210

ABSTRACT

There are natural geometric patterns in biology. Tissue layers, for example, differ mainly in the spatial distributions, size and packing of microstructure components such as the red blood cells, nuclei and cytoplasm etc. Expressive visualization by using the N-point correlation functions, involves the discovery of feature spaces that estimate and spatially delineate component distributions unique to a salient tissue. These functions provide feature spaces that are used to set useful transfer functions. We obtain insightful $3D$ visualizations of the epithelial cell lining in mouse mammary ducts and evolving structures in a zebrafish embryo. These are large datasets acquired from light and confocal microscopy scanners respectively.

Index Terms— microscopy imaging, N-point functions, tissue segmentation

1. INTRODUCTION

We wish to develop a general purpose methodology that uses geometric patterns and statistical distributions in the visualization of microscopy datasets. In biology, some of the structures are not immediately apparent at a microscopic scale. Figure 1(a) shows a typical image of a tissue employed in prototypical phenotyping studies [3]. Note the clear lack of a global $3D$ description of microscopic structures that populate each of the image planes. The cells on each plane are best identified by the presence of nuclei; however, the extent and proliferation of the nuclei in typical cellular constellations is not easily demarcated in $3D$. Further, tubular structures (e.g., ducts) which are prevalent in lesser numbers can only identified on each slide as a distinct projection of a complex tortuous $3D$ object.

In order to explore such structures, we need to determine their spatial locations of structures and enhance their visibility during scientific exploration. Spatial transfer functions available in the literature focus on structures that are in close proximity to other landmarks/surfaces. However, in biological data, salient structures are marked by a change in spatial distributions – which we intend to capture and hence naturally arrive at these structures when exploring. Note that the size of $3D$ datasets available in microscopy domains typically are quite large and manual searching is too cumbersome.

Furthermore, traditional transfer functions that are based on histogram features of the image intensities and gradients fails to resolve hidden $3D$ structures since the structural correlation across image planes is only partial. Figure 1(d) shows the rendering of a duct that we had to manually extract from the raw dataset. The lack of structural descriptions is evident in comparison to our automatically-detected-duct-renderings in (e) and (f). In general, capturing this structure and organization seems intractable owing to biological variations within and across samples. Biological structure in a soft tissue is deformable and seems to have non-linear structures with complex topology. The color luminance depends on acquisition parameters, staining etc on an image plane. There are image gradients within and across image planes. In addition, tremendous inter-dataset variations in shapes and size do not permit any standard methodology [5].

However, biological structure may be described in terms of physical cellular matrices. These matrices (of μm dimensions) include a coherent and mostly periodic collection of cells. At microscopic scales, tissue regions are formed from these matrices and effectively allow a hierarchical description of various larger structures. Hence, the question arises if there is a way to measure and differentiate microstructure ensembles. For example, in the H&E stained images from histology, a typical region consists of nuclei, red-blood cells (RBCs), cytoplasm and vacuous white space.

Microstructure, irrespective of its origin (material science or biology), may be measured as a collection (ensemble) of points, lines, internal surfaces, and volumes [1]. Each micro structural feature is associated with size, shape, volume, surface area, length and curvature attributes etc., morphological orientation, and location. Statistical distributions of such geometric attributes of ensembles collectively specify the geometric state of a microstructure. Mathematically, these properties of the microstructure are formalized by the statistical N-point correlation functions (N-pcfs) [7]. There are fundamental geometric constraints (such as polygonal separation of microstructure components) that are enforced while computing N-pcfs [2]. These constraints, therefore, provide useful

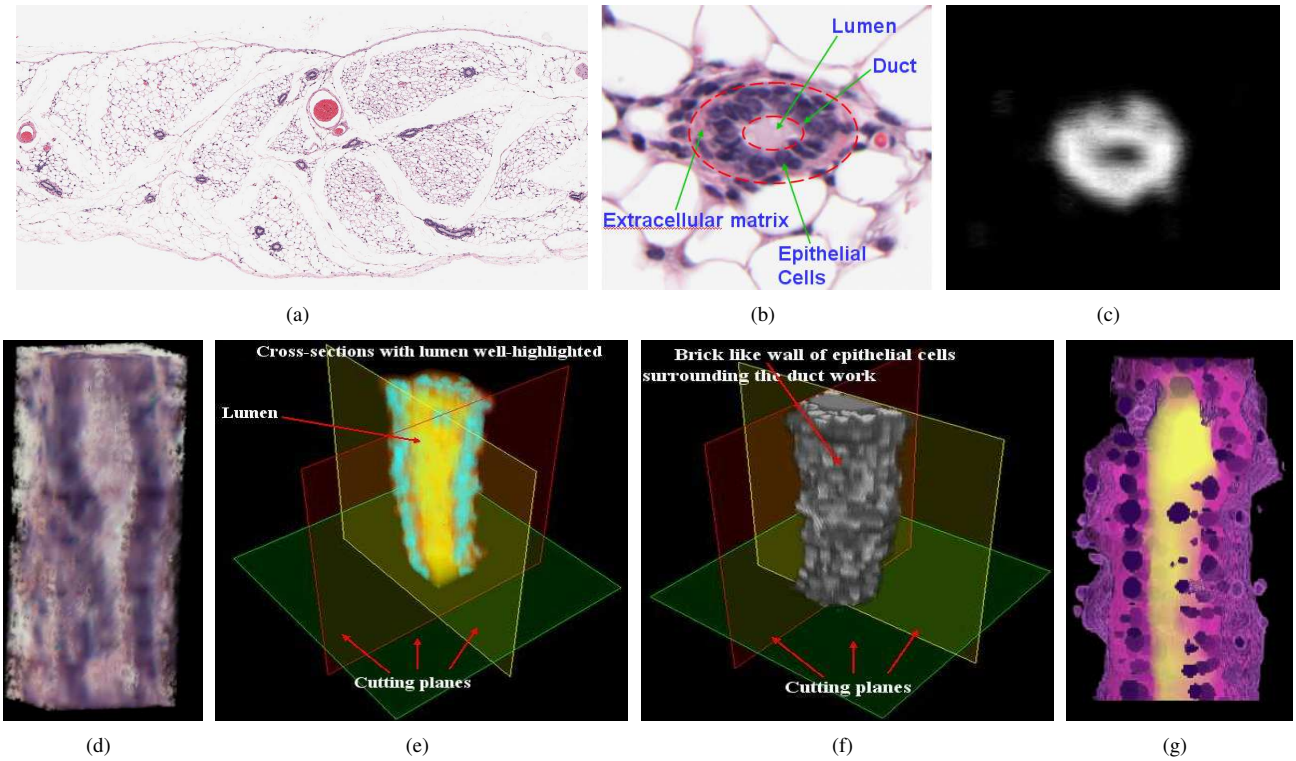


Fig. 1. (a) A histology section of the mouse mammary gland showing several duct cross-sections. (b) A zoomed duct section reveals the characteristic packing densities of the epithelial cell lining. (c) Color-mapped 2-pcf feature values. (d) Rendering of a duct from the raw data. The duct portion was manually cropped out of the large dataset. (e) *Categorical visualization* using the N-pcf feature values showing the chimney-like outer wall of the epithelial cells. (f) Axial section showing the lumen trapped in the duct. (g) An advanced visualization incorporating the 2-pcf feature space with nuclei segmentations.

user-input in choosing a certain functional form that is best representative of the microstructure.

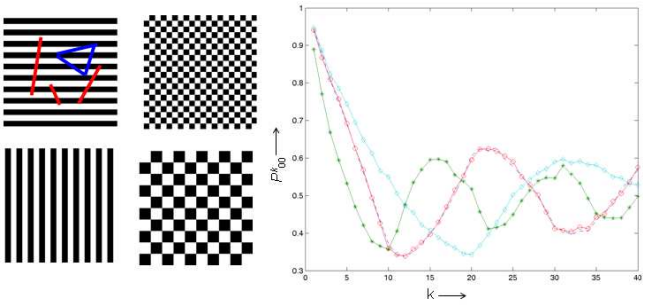


Fig. 2. The figure shows four different microstructure ensembles composed of 2-phase components, namely, phase 0 (black) and phase 1 (white). A plot of the orientation-averaged function P^k_{00} for varying k reveals different characteristic signatures.

2. N-POINT CORRELATION FUNCTIONS

To simplify the presentation, assume the presence of only two phases in the microstructure, namely, *phase 0* and *phase 1*. Let $p \in \Omega$ be any point in the microstructure $C : \Omega \rightarrow \{0,1\}$. Further, let $\Omega_p \subset \Omega$ be a suitable neighborhood of p. Consider placing a N-sided regular polyhedron with edge length k and orientation (θ,ϕ) in Ω_p. The probability that all the N-vertices lie in *phase 0* is defined as an N-point correlation function (N-pcf), $P^k_{i_1 i_2 \ldots i_N}(p,\theta,\phi)$, where $i_m = 0$ for $m \in 1,2,\ldots,N$. The subscript i_m denotes the phase of the m^{th} polyhedron vertex. The N-pcf for a regular polyhedron of edge length k depends on its orientation (θ,ϕ) and location p in the microstructure. In a similar manner, other forms of the point correlation function (with auto and cross-correlation phases at the vertices) may be defined. The orientation averaged N-*pcf* P^k_{ij}, which are of interest in this work, can be computed from the corresponding direction-dependent functions $\tilde{P}^k_{ij}(\theta,\phi)$ as follows:

$$< P^k_{i_1 i_2 \ldots i_N}(p) >= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \tilde{P}^k_{i_1 i_2 \ldots i_N}(p,\theta,\phi) d\theta d\phi \quad (1)$$

We now provide some insight into the probability measures captured by these functions. Consider the simple case of a 1-pcf, say P_0. It represents the probability that a point p is in *phase 0*. This quantity measures the volume fraction of *phase 0* in the microstructure. Similarly, P_1 is the volume fraction of *phase 1* and we have $P_0 + P_1 = 1$.

A 2-*pcf* is the probability of a straight line segment of length k randomly placed in the microstructure such that one end is in phase $i_1 \in \{0,1\}$ and the other end is in phase $i_2 \in \{0,1\}$. For a 2-phase microstructure, there are four possible 2-*pcf*s namely P^k_{00}, P^k_{01}, P^k_{10} and P^k_{11} and:

$$\begin{array}{lll} P^k_{00} + P^k_{01} + P^k_{10} + P^k_{11} = 1 & & P^k_{01} = P^k_{10}; \\ P^k_{00} + P^k_{01} = P_0; & & P^k_{10} + P^k_{11} = P_1 \end{array} \quad (2)$$

Please refer to Figure 2 wherein four different microstructures composed of black (0) and white (1) phases are considered. Note that each individual texture class in the image provides a unique or characteristic 2-*pcf* feature measure for a certain value of the separation distance k. Figure 2(right), a plot of the P_{11} measure for the four textures as a function of k is plotted. We observe that the four textures present characteristic signatures that enable their identification. For a given image, it is not known *a priori* what are the values of k. Hence, in practice, a range of values need to be explored while estimating these functions or alternatively, the given image needs to be inspected for suitable separation distances among components. The set of possible integral values that k may assume is represented by the discrete set $K \subset \mathbb{Z}$. The N-*pcf* feature descriptor for a tissue region represented by $P^{(k)}_{i_1 i_2 .. i_N} \in \Re^{K \times Q^N}$ is an $N+1$ mode tensor.

Essentially, a N-*pcf* is a multivariate distribution function. To estimate this function, we resort to using Monte Carlo sampling of the material components distribution using a sliding window (Ω_p). Conceptually, for a given separation length k, one needs to create auto- and cross-histograms at every point p in the discrete image plane. To estimate the functions at p, the number of samples (S) and window sizes (Ω_p) need to be specified. The minimum window size is proportional to the maximum separation distance that the functions will be evaluated for, and the sample size is chosen to keep the variance of the measured result to a desired range.

To evaluate the 2-*pcf* for separation distance k and phases i, j in a region containing m-phases, we place a number of randomly oriented and positioned lines. We then count the fraction of line segments that have an end point in phase i and the other in j to give an estimate of the 2-*pcf* P^k_{ij}.

3. RESULTS

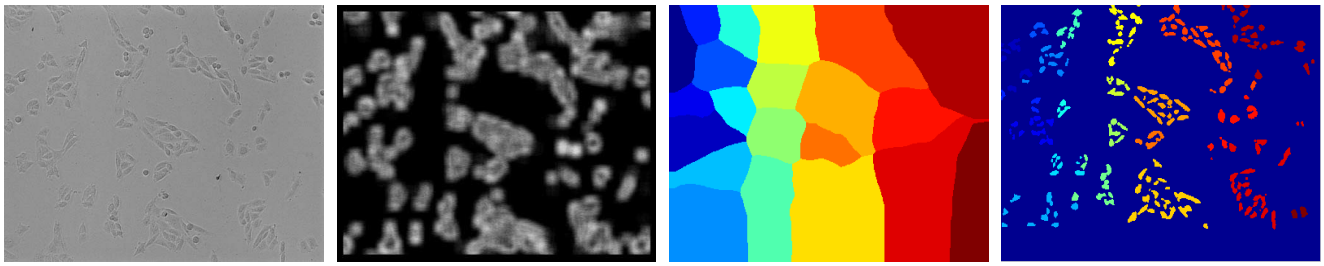
From Figure 1(b), we observe that the ducts have a concentric pattern of epithelial cells embedded in the extracellular material. The remainder of the tissue consists of fat cells with sparse populations of the nuclei and little or no extracellular material. Hence, we infer that a 2-pcf for the nuclei phase and extracellular material phase is sufficient to outline duct profiles. A stacking of the 2-pcf features computed on $2D$ images provided the following visualizations in $3D$. Figure 1(c) shows the generated feature values on the same mammary slice and the result after applying a threshold. Figure 1(e) shows the wall of the epithelial cell lining that is akin to a brick chimney wall. Figure 1(f) is an axial section of the duct revealing the lumen that lies trapped in the duct. Our visualizations allow biologists to see beyond the $2D$ planes that they are accustomed to. In phenotyping studies, genetic mutation leads to a change in structure and organization. Our techniques will allow the assessment of such changes quite naturally given our feature measurements.

A clonal colony consists of closely packed cells with low separation distance in the image domain. Using the cell segmentations in each frame, we group cells into existing clonal colonies. This requires us to develop a notion of distance of a cell from a clonal colony. The geodesic distance between a cell and the colony should be low if the cell belongs to the colony and vice-versa. Earlier, we mentioned about the similarity in cellular arrangements with material microstructure. Based on this observation, we utilize the N-point correlation functions (N-*pcfs*) to define a distance metric. These functions measure the cellular packing densities and evaluate to a high value in the interior of a colony region and diminish at the boundaries.

For this task, we are interested in temporally tracking the evolution of clonal colonies rather than individual cells. A segmentation of the individual cells is first achieved in each frame by using standard algorithms from literature [4]. We begin with a manual colony segmentation in the first frame. Based on the colony boundaries determined in the preceding frame, cells are grouped into colonies using a distance metric. Consequently, all the cells belonging to the same colony are grouped to obtain the current colony boundary.

We now describe the distance metric that we compute [6]. Within a colony region, P_{11} (where 0 represents the background and 1 represents the nuclei phase) has a high value (close to 1) than elsewhere (Figure 3a). Therefore, a geodesic distance metric $G = \nabla P_{11} \nabla P^T_{11}$ is defined on the 2-*pcf* manifold to measure proximity during tracking (Figure 3b).

We reason that the performance of our algorithm depends on the 2-*pcf* in effectively defining a colony region. In order to validate our algorithm, consecutive frames were reviewed and all pertinent colonies were delineated with boundaries. We compared the algorithm-detected colony cells and the ground-truth colony cells. The matched pairs were separated and counted in 4 categories: (i) a - *true-positives* (ii) b - *false-positives* (iii) c - *false-negatives* (iv) d - *true-negatives*. Thus, one can define the sensitivity $S = \frac{a}{a+c}$, specificity $Sp = \frac{d}{b+d}$ and error $E = \frac{b+c}{a+b+c+d}$. Our results are presented on five datasets in Table 1. The first 3 entries are for the three frames



(a) (b) (c) (d)

Fig. 3. (a) Phase-contrast microscopy images of mutated cell colonies that are rapidly proliferating in culture. Each colony is marked by a dense packing of cells, (b) 2-*pcf* evaluated on the first frame shown in (a), (c) Voronoi segmentation on the 2-*pcf* manifold based on colony boundaries of the preceding frame (c) Colony segmentations.

of a single dataset and the last entry averages the parameters for the remaining four datasets.

Frame	Cell Counts	S	Sp	E
Frame 1	378	0.9134	0.8827	0.0215
Frame 2	610	0.8407	0.8222	0.123
Frame 3	794	0.7807	0.8022	0.1733
Average	-	0.8417	0.8322	0.1033

Table 1. Sensitivity/specificity values for colony tracking

We observe that the framework has a good sensitivity ($>$ 0.84), specificity ($>$ 0.83) and with a low error rate ($<$ 0.15) on average. It is easy to see that the performance of the algorithm is best in the earlier frames and relatively diminishes in later frames. This is expected since the extensive clonal growth creates merged boundaries. In turn, this leads to the cells lying on the Voronoi boundaries to be easily affected. We also observed in several instances that a single colony in the starting frame emerged as two separate colonies in later frames. As a result of which the tracking step combined the two colonies into one of the same affiliation.

4. SUMMARY AND FUTURE WORK

In this paper, we described a tissue visualization algorithm that is applicable to optical microscopy data. We estimate the packing and material component distributions locally using the N-point correlation functions. These functions are realized using suitable windowing and sampling strategies to provide feature representations. Our methods have been applied in two phenotyping studies involving the mouse mammary ducts and zebrafish embyrogenesis. In future, we would like to rigorously evaluate the N-pcf features towards setting useful transfer functions that enhance the expressivity of the visualization process.

5. REFERENCES

[1] A. Gokhale. Experimental measurements and interpretation of microstructural n-point correlation functions. *Microscopy and Microanalysis*, 10:736–737, 2004.

[2] A. Gokhale, A. Tewari, and H. Garmestani. Constraints on microstructural two-point correlation functions. *Scripta Materialia*, 53:989–993, 2005.

[3] K. Kurose, K. Gilley, S. Matsumoto, P. Watson, X. Zhou, and C. Eng. Frequent somatic mutations in PTEN and tp53 are mutually exclusive in the stroma of breast carcinomas. *Nature Genetics*, 32(3):355–357, 2002.

[4] K. Mosaliganti, L. Cooper, R. Machiraju, K. Huang, and G. Leone. Reconstruction of cellular biological structures from optical microscopy data. *IEEE Transactions on Visualization and Computer Graphics (in review)*, 2007.

[5] K. Mosaliganti, F. Janoos, O. Irfanoglu, R. Ridgway, R. Machiraju, K. Huang, J. Saltz, G. Leone, and M. Ostrowski. Tensor classification of n-point correlation function features for histology tissue segmentation. *Submitted to the Journal of Medical Image Analysis*, 2007.

[6] K. Mosaliganti, X. Xu, R. Machiraju, K. Huang, and W. Xia. Automated quantification of colony growth in clonogenic assays. In *MICCAI Workshop on Medical Image Analysis with Applications in Biology*, 2007.

[7] S. Torquato. *Random Heterogenous Material*. Springer Verlag, 2004.